The Astrophysical Journal, 712:130–141, 2010 March 20 doi:10.1088/0004-637X/712/1/130

DARK FLUID: A UNIFIED FRAMEWORK FOR MODIFIED NEWTONIAN DYNAMICS, DARK MATTER, AND DARK ENERGY

HongSheng Zhao[1] and Baojiu Li[2]
[1] Scottish University Physics Alliance, University of St. Andrews, KY16 9SS, UK; hz4@st-andrews.ac.uk
[2] Department of Applied Math and Theoretical Physics, Cambridge University, CB3 0WA, UK; b.li@damtp.cam.ac.uk
Received 2008 April 7; accepted 2009 December 7; published 2010 February 25

ABSTRACT

Empirical theories of dark matter (DM) like modified Newtonian dynamics (MOND) gravity and of dark energy (DE) like $f(R)$ gravity were motivated by astronomical data. But could these theories be branches rooted from a more general and hence generic framework? Here we propose a very generic Lagrangian of such a framework based on simple dimensional analysis and covariant symmetry requirements, and explore various outcomes in a top–down fashion. The desired effects of quintessence plus cold DM particle fields or MOND-like scalar field(s) are shown to be largely achievable by one vector field only. Our framework preserves the covariant formulation of general relativity, but allows the expanding physical metric to be bent by a single new species of dark fluid flowing in spacetime. Its non-uniform stress tensor and current vector are simple functions of a vector field with variable norm, not coupled with the baryonic fluid and the four-vector potential of the photon fluid. The dark fluid framework generically branches into a continuous spectrum of theories with DE and DM effects, including the $f(R)$ gravity, tensor–vector–scalar-like theories, Einstein–Aether, and $\nu\Lambda$ theories as limiting cases. When the vector field degenerates into a pure scalar field, we obtain the physics for quintessence. Choices of parameters can be made to pass Big Bang nucleosynthesis, parameterized post-Newtonian, and causality constraints. In this broad setting we emphasize the non-constant dynamical field behind the cosmological constant effect, and highlight plausible corrections beyond the classical MOND predictions.

Key words: cosmology: theory – dark matter – galaxies: kinematics and dynamics – gravitation

Online-only material: color figures

1. INTRODUCTION

Gravity, the earliest and the weakest of the known forces, has never been very settled. The beauty of covariant symmetry motivated Einstein to supersede Newton's paradigm with general relativity (GR), but (inadequate) empirical evidence motivated Einstein to introduce first and then withdraw the cosmological constant, a concept defying quantum physics understanding even in modern day. While making generally no more than a factor of 2 corrections to Newton's theory, GR and its equivalence principles insist on covariant symmetries in spacetime. There is also no frame to measure locally any absolute direction of gravitational acceleration for a free-falling observer. However, symmetry can be spontaneously broken if there are dynamical interactions or couplings of fields, a well-known mechanism in several branches of physics, especially the Higgs mechanism in particle physics that gives a mass to a particle. Many attempts have been made to break the strong equivalence principle by adding new fields (degrees of freedom) in the gravity sector, which essentially means the gravitational "constant" G may be a new dynamical degree of freedom governed by other fields coupled to the metric. The best known is the Brans–Dicke theory (Brans & Dicke 1961). The lesser known is that a vector field of a non-zero absolute value in vacuum can also be coupled to gravity, to give absolute directions (Will 1993). It has long been suggested that Lorentz symmetry can be broken locally in the quantum gravity and string theory context (Kostelecky & Samuel 1989) to yield a vector field of a non-zero expectation value (e.g., pointing toward the direction of time) in the vacuum. The most successful attempt so far is the Einstein–Aether theory of Jacobson & Mattingly (2001). A common theme of these theories is that they are *not* invented for certain observational anomaly. Rather in the same vein as how symmetry motivated GR, these theories meet the astronomical data only a posteri, e.g., Li et al. (2008) showed a vector field in the gravity sector could *not* be excluded by the accurate cosmic microwave background (CMB) data.

Nevertheless, the above order is not the only way to discover theories. The puzzling blackbody radiation spectrum and Balmer's curious empirical formula for hydrogen lines are among the odd pieces of classical physics that led to the full formulation of quantum mechanics. This bottom–up approach is often gradual, the arrival of the final theory taking several generations of formulations (e.g., from Planck's model for blackbody radiation and Bohr's model for a hydrogen atom to Heisenberg's matrices-based formulation in general) with different levels of mathematical rigor and sophistication.

Historically, Milgrom's modified Newtonian dynamics (MOND) was invented without any packaging by covariant theories of gravity, just as the concept of dark matter (DM) was invented by Zwicky without packaging first with supersymmetry-like particle field theories. MOND is a model motivated to explain the curious uniform rules (or facts) underlying rotation curves of many spiral galaxies, as Balmer's formula and its generalizations suggesting strongly a fundamental rule for all atomic lines. Since the rule is empirical and bare (without covariance), it waits to be enshrouded by a theory preserving basic symmetries to predict any logical corrections to situations where the empirical rule must fail slightly, e.g., by a factor of 2 in some gravitational lenses made by elliptical galaxies and clusters of galaxies. The tensor–vector–scalar (TeVeS) framework of Bekenstein (2004), building up earlier constructions (e.g., Bekenstein 1988; Bekenstein & Sanders 1994) and especially the introduction of a vector field (Sanders 1997), makes the first step to the integration of the MOND formula with fundamental physics. A time-like vector field is shown to be the necessary ingredient of a MOND gravity. However, the vector field does not completely replace the role of the CDM particles (Zuntz et al. 2008) or neutrinos (Skordis et al. 2006), even though it is capable of driving some structure growth (Dodelson & Liguori 2006). Yet the original aim of TeVeS was limited, e.g., it is neither a model for dark energy (DE) nor for structure formation. Orthogonally many literatures considered theories of modified gravity such as the $f(R)$ gravity (Chiba 2003) and scalar inflation theory as ad hoc fixes of the cosmological constant problem and the horizon problem, respectively, without aiming to address outstanding questions on galaxy rotation curves. Recently, Zhao (2007) and Halle et al. (2008), building on the work of Zlosnik et al. (2007), showed that these outstanding problems of DM and DE can find at least one common solution simultaneously in the framework of a massive vector field, which is called a $\nu\Lambda$ model. In these theories, there is "one field which rules them all and in the darkness bind them." These models share as inspirations models like Chapligin gas (e.g., Bento et al. 2002) and other unified models (e.g., Peebles & Vilenkin 1999), but use the Tully–Fisher relations explicitly as the basis of the models.

The goal of the paper is to build a framework of new kinds of models. We shall show that MOND, $f(R)$ gravity, Einstein–Aether theory, and DE theories can be integrated into a common covariant framework with a Lagrangian depending on a unit vector and a dynamical scalar field.[3] MOND would become a specific choice of the potential of the scalar field. Having such a framework allows one to explore the consequences of modified gravity systematically. It can be meaningless to even differentiate DE and modified gravity. Modified gravity contains extra fields, which can be treated as a DE field.

One of the goals of the paper is to show covariant corrections to the MOND formulae for a time-dependent system. There can also be corrections at very small spatial scale. These are the generic consequences of Lagrange equations.

The outline is as follows. We propose our general Lagrangian in Section 2, and choose a subset of models with MOND and DE effects. We discuss the properties of our dark fluid in the cases of Hubble expansion (Section 3), and for static galaxies (Section 4), where we give the modified Poisson equation and the equation for the Hubble expansion. We discuss corrections to MOND in Section 5. We summarize the properties of the dark fluid in Section 6. In Appendices A–E we give the full Lagrangian, full equations of motion (EOMs) of the theory, an estimate of the damping frequency of the dark fluid, and a choice for the fluid's potential function, and some background information of the 3+1 decomposition formulation.

2. A SIMPLE LAGRANGIAN FOR THE DARK FLUID

We propose a Lagrangian $\mathcal{L}$ containing a scalar φ and a time-like unit vector $Æ^a$, and the metric tensor g_{ab} as

$$\mathcal{L} = R - \frac{c_\varphi^2}{N^2}(\nabla_c\varphi\nabla^c\varphi) + 2\Lambda_{\rm int} + \mathcal{L}_{Æ} + \not{I}, \qquad (1)$$

where N, c_φ^2 are various coupling constants or functions of $\lambda = \varphi^2$, R is the Ricci scalar, and $\Lambda_{\rm int}$ takes the role of the scalar potential or interaction or the mass term:[4]

$$\Lambda_{\rm int}(\varphi, K_4) = \Lambda_0 F(\varphi^2) + \varphi^2 K_4, \quad K_4 \equiv \nabla_\parallel Æ_c \nabla_\parallel Æ^c, \qquad (2)$$

where $\Lambda_0 \equiv (1.2 \times 10^{-10} \text{ m s}^{-1})^2$ and K_4 are important for creating the MOND effect, and $K_4 = 0$ for a uniform cosmology; the notation $\nabla_\parallel \equiv Æ^a\nabla_a$ is a derivative parallel to the local time direction, and $\nabla^a = g^{ac}\nabla_c$, $Æ_a = g_{ac}Æ^c$. The term $\Lambda_{\rm int}$ takes on several roles: it gives an effective potential or mass for the scalar field, and allows the mass or sound speed to vary with the interaction of φ and Æ; to achieve the cold dark matter (CDM) effect during structure formation would require $\varphi^{-2}\Lambda_{\rm int} \to \Lambda_0\varphi^{-2}F \to cst$ at high redshift when $\varphi \to \infty$, i.e., a nearly constant scale-free mass. To achieve the cosmological constant effect would require that $\Lambda_{\rm int} \to \Lambda_0 F \to cst$ in the future when $\varphi \to 0$.

Other interaction terms are collected in

$$\not{I} = e_0\varphi^2 R + [c_1\nabla_a Æ_c + c_3\nabla_c Æ_a - (e_1 Æ_a g_{bc} + e_2 Æ_b g_{ac})\nabla^b\varphi]\nabla^a Æ^c, \qquad (3)$$

$$\mathcal{L}_{Æ} = \left[c_2(\nabla_a Æ^a)^2 - \frac{1-c_\varphi^2}{N^2}(\nabla_\parallel\varphi)^2\right] + (g_{ab}Æ^a Æ^b - 1)L^*, \qquad (4)$$

where L^* is the Lagrangian multiplier.

In an *effective* sense, we can pack our $Æ^a$ and φ fields as one field: our Lagrangian can be cast in terms of a vector field of dynamics norm $Z^a \equiv Æ^a\varphi$ where one can pack up the 3+1 degrees of freedom into a vector of 4 degrees of freedom in the same style as in Halle et al. (2008). The implications of such a vector field with a norm $\lambda \equiv \varphi^2$ is discussed in Appendix A, where we illustrate how a TeVeS Lagrangian $L(Z^a) = L(\varphi Æ^a)$ reduces to various special cases, TeVeS, BSTV, Einstein–Aether, $f(K)$, $f(R)$ (e.g., Carroll & Lim 2004; Lim 2005; Sanders 2005; Kanno & Soda 2006).

For a first study, we shall set

$$\not{I} = 0 \qquad (5)$$

hereafter by setting the coefficients c_1, c_3, e_0, e_1, e_2 (which are generally functions of φ) to zero. Specifically, we choose to set $e_0(\varphi) = 0$ to eliminate $f(R)$-gravity-like coupling and set the functions $e_1(\varphi) = e_2(\varphi) = 0$ to eliminate kinetic cross-coupling like $\nabla Æ\nabla\varphi$, and $c_1 = c_3 = 0$ to eliminate the spin-1 mode gravitational wave. This is done partly to simplify the dynamics, partly because any complicated couplings can be dangerous theoretically (see Appendix B); these coupling terms serve as an effective potential for the scalar φ, and this dynamically changing potential does not appear to have a well-defined lower bound to safeguard the causal structure. It is tempting to set $c_2 = 0 = 1 - c_\varphi^2$ for simplicity; however, we will not do so yet since these terms in $\mathcal{L}_{Æ}$ could be desirable for a theory to be healthy.

2.1. Parameters for MOND-inspired Subset of Dark Fluid Theories

A theory is built by minimizing the action $S = -(16\pi G)^{-1}\int dx^4\sqrt{-g}(\mathcal{L}_m + \mathcal{L})$ once the Lagrangian density

[3] A possible origin for the dark fluid is a field coupled to neutrinos of various flavors; this field allows transitions of neutrinos of various energy, mass, flavor, and helicity, and there are many possibilities within theories of neutrino mass (Zhao 2008a).

[4] Other functions $\Lambda_{\rm int} = f(y)(\varphi^2+1)\Lambda_0$ with $y = \frac{K_4}{(1+\varphi^2)\Lambda_0}$ for some function f can also create MOND, DM, and cosmological constant.

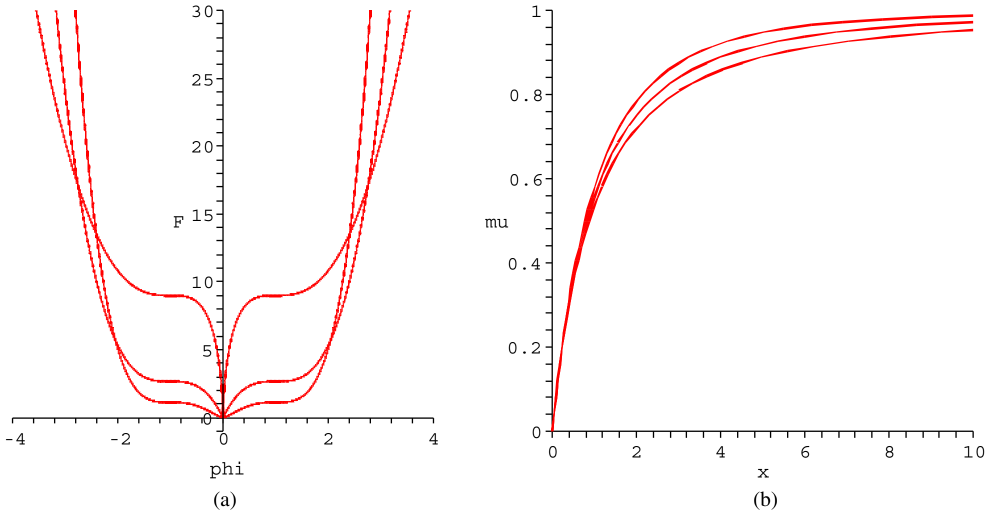


Figure 1. Panel (a): $F(\varphi^2)$ vs. φ for $k = 3, 3/2, 2$ (top to bottom); note the symmetric potential with the global minimum $F = 0$ at $\varphi = 0$, and the plateau at $\lambda = 1$. Panel (b) shows our function $\mu \equiv 1 - \varphi^2$ as function of $x \equiv |\nabla\Phi|/\sqrt{\Lambda_0}$ for the models with $k = 3, 3/2, 2$ (top to down), adopting $\sqrt{\Lambda_0} \to a_0$.
(A color version of this figure is available in the online journal.)

$\mathcal{L}$ is specified. Our Lagrangian $\mathcal{L}$ has three dynamical fields: the scalar field λ (or equivalently φ), the Æther field $Æ^a$, and the metric field g^{ab}. Now minimizing the action S with respect to variations of the three fields will lead to the scalar field EOM, Æther field EOM, and the modified energy–momentum tensor. Variations with the non-dynamical Lagrange multiplier L^* gives the unit-vector constraint for the Æ field. The general results are more tedious and are presented elsewhere (Halle et al. 2008). To illustrate the physics and stay away from models with obvious pathological dynamics, we consider only a specific choice of functions to select a MOND-related subset of dark fluid theories. Starting from the Lagrangian given in Equation (1), we select models with $\not{L} = 0$ and select the functions in $\mathcal{L}_\varphi$ and $\mathcal{L}_{Æ}$ as follows.

The dimensionless function $F(\lambda)$ has the meaning of the potential of the scalar field $\lambda \equiv \varphi^2$. The coupling constants c_φ^2 and c_2 are of order unity, and will be shown to be related to the sound speeds of the dark fluid.

1. Our choice for the dark fluid potential (shown in Figure 1) with a constant energy scale Λ_0 is

$$\Lambda_0 F(\varphi^2) = \frac{k^3\Lambda_0}{3} + \frac{k^3\Lambda_0}{3}(|\lambda_\pm|^{1/k} - 1)^3, \quad k = 3, \tag{6}$$

where we shall adopt $\lambda_\pm \approx \lambda = \varphi^2$ essentially up to a small correction.[5] The reason to choose $k = 3$ is such that we make the potential $\Lambda_0 F \to 9\Lambda_0\varphi^2$ to be harmonic at large φ to mimic CDM.[6] In galaxies we can neglect ϵ; so $\lambda_\pm \approx \lambda \equiv \varphi^2$ and $F = 9(\varphi^{2/3} - 1)^3 + 9 = 9(\varphi^2 - 3\varphi^{4/3} + 3\varphi^{2/3})$. An important asymptotic property of the functions F and F' (a prime $'$ always means a derivative $\frac{d}{d\lambda}$) is that

$$F' \equiv \frac{d}{d\lambda}F \sim (1 - \lambda^{-1/3})^2, \quad \text{if } \lambda \equiv \varphi^2 \to 1, \tag{7}$$

$$\propto (\lambda - \epsilon^2)^{-1/2}, \quad \text{if } \lambda \equiv \varphi^2 \to +\epsilon^2 \sim 10^{-7}, \tag{8}$$

$$\sim 0 \quad \text{if } \lambda \equiv \varphi^2 \to \infty. \tag{9}$$

We shall show that this property describes a non-uniform (DE) fluid which gives the MOND-like (DM) effects in galaxies, and Newtonian-like effects in the solar system.

2. We set the scalar field sound speed as a finite positive constant, i.e.,

$$\infty > c_\varphi > 0, \tag{10}$$

and we shall treat N as constants too. In general, theories with finite positive propagation speeds have a finite causality cone, hence are well behaved (Bruneton & Esposito-Farese 2007). The opening angle of the cone of propagation can be wider than that of photons.

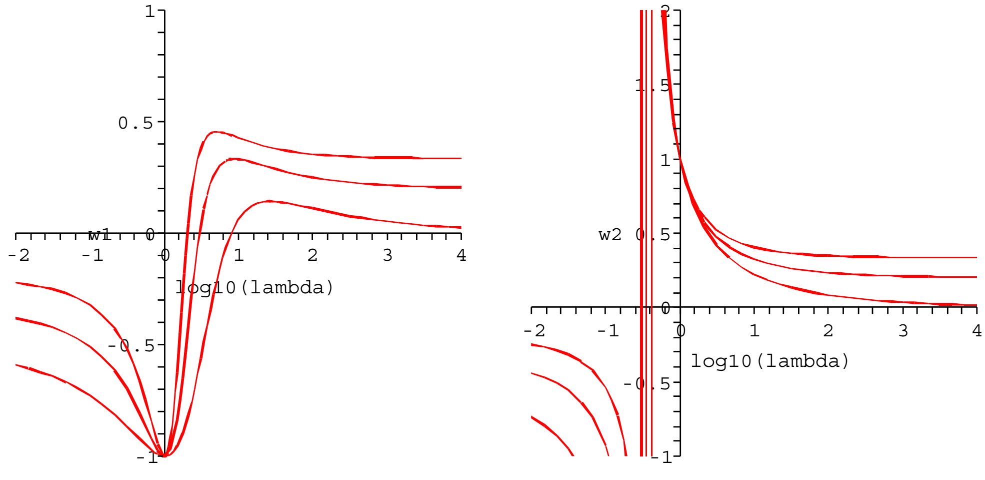
3. We set the equivalent part of Jacobson's unit vector field Lagrangian to be $c_1K_1 + c_3K_3 + c_4K_4 + c_2K_2 = 0 \times K_1 + 0 \times K_3 + 2\lambda K_4 + c_2K_2$, where, e.g., c_2K_2 corresponds to the second term in $\mathcal{L}_{Æ}$. This choice $c_1 = c_3 = 0$ in the Lagrangian kills spin-1 mode waves of the vector field, and guarantees that the normal gravitational wave in the tensor mode will propagate with the normal speed (of light). This might not be necessary, but it simplifies the analysis of parameterized post-Newtonian (PPN) parameters in the solar system and the sound speed of the vector field.

One can use the analysis by Foster & Jacobson (2006) to predict the sound speed square of the spin-0 mode of Aether fluctuation, $c_0^2 \equiv \frac{(c_2+c_1+c_3)}{(c_4+c_1)(1-c_1-c_3)}\frac{(2-c_4-c_1)}{(2+c_1+3c_2+c_3)} = \frac{(1-\lambda)c_2}{(2+3c_2)\lambda}$. Among the many choices[7] of c_2 to guarantee that c_0^2 and c_2

[5] More rigorously we make a small correction, e.g., $\lambda_\pm = \frac{\varphi^2+\epsilon^2}{1+\epsilon^2}$ or $\lambda_\pm = \frac{\varphi^2 \pm \varphi^{-2}\epsilon^4}{1-\epsilon^4}$ with $\epsilon^2 \sim O(10^{-7})$ being a small positive constant, which forms a barrier preventing the theory from entering unphysical regime in the solar system (cf. Appendix C). Note our choice of the function $\Lambda_0 F(\varphi^2) \approx k^3\Lambda_0[\frac{1}{3}\lambda^{3/k} - \lambda^{2/k} + \lambda^{1/k}]$ has the zero point $\Lambda_0 F|_{\lambda=\varphi^2=0} = 0$, i.e., it is designed to give no cosmological constant like effect in the solar system. At the other extreme $\Lambda_0 F(1) = \Lambda_0 k^3/3$, hence creating a constant potential at the cosmological scale.

[6] Other choices are also interesting: a model with $k = 3/2$ would represent relativistic DM at large φ with $\Lambda_0 F \to \frac{9}{8}\Lambda_0\varphi^4$. A model with $k > 3$ could drive inflation with an appropriate N. A model with $k = 2$ is the simplest with $\Lambda_0 F = \frac{8\Lambda_0}{3}[(|\varphi| - 1)^3 + 1] = 8\Lambda_0(\frac{1}{3}|\varphi|^3 - |\varphi|^2 + |\varphi|)$.

[7] For example, a simple choice for c_2 is $c_2 = \frac{2}{3N^2}\frac{1-\lambda}{1+\lambda}$. This guarantees that $c_0^2 = \frac{(1-\lambda)^2}{3N^2(1+\lambda)\lambda} \geqslant 0$; a large N can guarantee that the effect of the c_2K_2 term is small since $|c_2K_2| \leqslant \frac{2K_2}{3N^2} \sim \frac{2H^2}{N^2}$, although it is unclear if it is desirable that $c_0^2 = 0$ at $\lambda = 1$, and c_0^2 can be small in the solar system.



(a) (b)

Figure 2. Panel (a): equation of state estimator $w_1 = \frac{\varphi^2 F' - F(\varphi^2)}{\varphi^2 F' + F(\varphi^2)}$ vs. $\log_{10}(\lambda)$ for $k = 3, 3/2, 2$ (bottom to top); note that w_1 becomes negative for $\lambda < 10$. Panel (b) shows the sound of speed estimator $w_2 = \frac{\delta[\varphi^2 F' - F(\varphi^2)]}{\delta[\varphi^2 F' + F(\varphi^2)]}$. Note that $k = 3$ has very small sound speed a large λ, but there is a singularity near $\lambda \sim 1/3$.
(A color version of this figure is available in the online journal.)

are smooth functions for all values of λ, the simplest choice is such that

$$-\frac{c_2}{2} = \frac{1}{3\lambda} \quad \rightarrow \quad c_0^2 = \frac{1}{3\lambda}. \tag{11}$$

The important thing is that c_0^2 is a finite real positive number given by $c_0^2 = \frac{1}{3\lambda} > 0$. Interestingly, in the solar system, where the scalar field λ is expected to settle to a very small equilibrium value $\lambda \sim \epsilon^2 \sim O(10^{-7})$ for our choice of the penalizing scalar field potential $F(\lambda)$, the spin-0 mode of the vector field propagates with a finite superluminal speed, $c_0 = (3\epsilon^2)^{-1/2} \gg 1$, avoiding the Cherenkov radiations constraint in the solar system. All PPN parameters are expect to be equal to that of GR in the solar system as well; the PPN parameters $\alpha_1 = -8\lambda$, $\alpha_2 = (3\lambda - 1)\lambda$ are non-zero, but can be made small enough to fit current limits $|\alpha_1| < 10^{-4}$ and $|\alpha_2| < 4 \times 10^{-7}$ (Foster & Jacobson 2006).

The general EOMs and Einstein equations are given in Appendix A. In the following section, we will apply these to simple configurations: uniform cosmology and static galaxies.

3. BACKGROUND COSMOLOGY

Consider background cosmology in the FRW flat metric,

$$ds^2 = dt^2 - a^2(t)(dx^2 + dy^2 + dz^2). \tag{12}$$

First, the scalar field follows an EOM exactly as a quintessence (see Equation (C2)),

$$\ddot{\varphi} + 3H\dot{\varphi} = -\left(\Lambda_0 F' + \frac{3\alpha'}{2} H^2\right)(2N^2\varphi), \tag{13}$$

so φ tracks the Hubble rate $H = \dot{a}/a$. Here we introduce an auxiliary function $b(\lambda)$ or $\alpha(\lambda)$ defined by $b(\lambda) - 1 \equiv \frac{\alpha(\lambda)}{2} \equiv \frac{c_1 + c_3 + 3c_2}{2} = -\lambda^{-1}$ for our choice of coefficients $c_2 = -\frac{1}{3\lambda}$, where $\lambda = \varphi^2$. The vector field EOM gives the Lagrange multiplier $L^* = \partial_t(\alpha H) + \frac{1 - c_\varphi^2}{N^2}\dot{\varphi}^2$. Since L^* and $N^2\Lambda_0 F'$ can be thought as the masses of the scalar and vector fields, hence our model describes effectively an unstable slowly decaying dark particle. To be more explicit, consider the approximation $\alpha' \rightarrow 0$ and $\Lambda_0 F \sim 9\varphi^2$ for large φ in our model with $k = 3$. The scalar EOM becomes a damped harmonic equation,

$$\ddot{\varphi} + 3H\dot{\varphi} = -18N^2\Lambda_0\varphi, \tag{14}$$

which gives the approximate solution

$$9\Lambda_0\varphi^2 \approx 3\Omega_0 H^2 \cos^2(Nt\sqrt{18\Lambda_0}),$$
$$\frac{1}{2N^2}\dot{\varphi}^2 \approx 3\Omega_0 H^2 \sin^2(Nt\sqrt{18\Lambda_0}), \tag{15}$$

which is rigorous in the matter-dominated era when $H^{-1} = \frac{3}{2}t$, where $\Omega_0 \sim O(1)$ is a constant, determined by the initial condition of φ. Taking the average over the rapid oscillations, we have

$$(F - 9)\Lambda_0 \approx \frac{1}{2N^2}\dot{\varphi}^2 \approx \frac{3\Omega_0}{2} H^2 \tag{16}$$

at high redshift. A numerical example of the oscillating scalar field is given in Figure 3.

The modified 00 term of the Einstein equation becomes (see Equation (C8))

$$3H^2 = 8\pi G\,[\bar{\rho} + \bar{\rho}_{\rm DS}]\,, \tag{17}$$

$$\bar{\rho}_{\rm DS} \equiv \frac{1}{8\pi G\left(1 + \frac{\alpha}{2}\right)}\left[\frac{1}{2N^2}\dot{\varphi}^2 + \Lambda_0 F - \alpha 4\pi G\bar{\rho}\right] \tag{18}$$

where $\bar{\rho}_{\rm DS}$ and $\bar{\rho}$ are the (background) energy densities of the dark sector (DM plus DE) and the baryon-radiation fluid, respectively. At high redshift when $F \rightarrow \lambda$ we get a DM-like effect, and $[\frac{1}{2N^2}\dot{\varphi}^2 + \Lambda_0\varphi^2]$ behaves like a CDM of the Broglie frequency $\sqrt{2N^2\Lambda_0}$. The DM-like effect is sub-dominant at very high redshift, e.g., radiation era or Big Bang nucleosynthesis (BBN), where $\lambda \rightarrow \infty$, hence $\alpha \rightarrow 0$. Therefore, the BBN constraint is automatically satisfied because the Hubble expansion

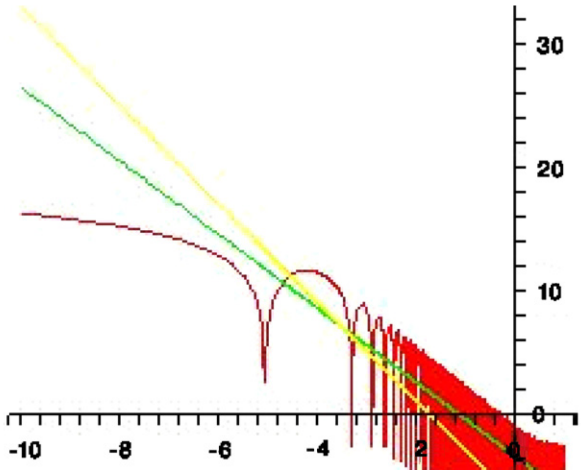


Figure 3. Contributions to the Hubble expansion $\ln(H^2/H_0^2)$ from three components as a function of the scale factor $\ln a$: baryonic matter (green straight line with an a^{-3} fall-off), radiation (yellow straight line with an a^{-4} fall-off), and dark fluid's potential term $a_0^2 F/3$ (the oscillating curve with a nearly a^{-3} fall-off of its amplitude for the epoch between $1 \gg H_0 t \gg 1/N$). For illustration purposes, we assume parameters $\Omega_b = 10\,\Omega_r = 0.03$, and $N = 1000$, an initial condition $\phi = 1.8N$ at $\ln a = -10$. Note the dark fluid is sub-dominate in the radiation era, but dominates the baryonic fluid by a factor of 2 in the matter era, and starts to flat out to a constant amplitude at the present epoch, similar to the cosmological constant effect. More realistic models with a larger N are challenging to integrate numerically (appropriate for studying inflation and DM effects in galaxies), and the results are sensitive to the initial condition of the scalar field. Nevertheless, the dark fluid seems to have the desired properties of in between that of a uniform cosmological constant and that of a CDM.

(A color version of this figure is available in the online journal.)

rate at BBN is equal to that of a radiation-only universe. We also expect the early structure formation to resemble CDM models.

Equivalently, the Einstein equation can be written as

$$-\left(2\frac{\ddot a}{a} + H^2\right) = 8\pi G\,(\bar p + \bar p_{\rm DS}) \tag{19}$$

$$\bar p_{\rm DS} \equiv \frac{1}{8\pi G\left(1+\frac{\alpha}{2}\right)}\left[\frac{1}{2N^2}\dot\varphi^2 - \Lambda_0 F + 2\alpha' H\varphi\dot\varphi - \alpha 4\pi G\bar p\right], \tag{20}$$

where the pressure of the baryon-radiation fluid $\bar p = 0$ in the matter-dominated era. Applying the harmonic approximation $\frac{1}{2N^2}\dot\varphi^2 \sim \Lambda_0[F - F(1)]$ we find the effective pressure of the dark sector

$$-\bar p_{\rm DS} \sim \frac{9\Lambda_0}{8\pi G\left(1+\frac{\alpha}{2}\right)}. \tag{21}$$

This behaves like a DE with a characteristic scale Λ_0, which is of the same order of magnitude (smaller by a factor of a few)[8] as the observed cosmological constant $\sim (8\times 10^{-10}\ \mathrm{m\ s^{-2}})^2$. Note that in writing the above equations we have implicitly assumed that the effective DM and effective DE components couple to each other. This can been seen by checking that neither DM nor DE satisfies the conservation law $\dot{\bar\rho} + 3H(\bar\rho + \bar p) = 0$, but their sum $\bar\rho_{\rm DS}$ does. Figure 2 show two crude estimators for the equation of state parameter and possible problematic regions.

[8] However, an interesting choice for $\Lambda_{\rm int} = \frac{n^3}{3}\Lambda_0(1+\theta) + (1+\theta^{1/3})^{2n}(1-n\theta^{1/3})K_4$, for $\theta \equiv (\lambda^{1/3}-1)^3$, especially $n = 6$ to produce the right amount of cosmological constant $n^3\Lambda_0/3 \sim 72\Lambda_0$.

4. STATIC GALAXY LIMIT

To work out perturbations in static galaxies, remember that in the Newtonian gauge we have only two scalar mode perturbation potentials, Φ and Ψ, which appear in the perturbed metric:

$$ds^2 = (1+2\Phi)dt^2 - a_0^2(1+2\Psi)(dx^2+dy^2+dz^2), \tag{22}$$

where we will let $a_0 = 1$. As a first study we assume no Hubble expansion.

The vector field EOM in static systems fixes $Æ_a = g_{ab}Æ^b = (1+\Phi, 0, 0, 0)$, so the vector field tracks the metric exactly without any freedom in static galaxies.

The 00 component of the Einstein equation becomes a Poisson equation (see Equation (C8)),

$$\sum_{i=x,y,z} -(2\Psi)_{,ii} = 8\pi G(\rho + \rho_{\rm DM} + b\rho_{\rm DE}), \tag{23}$$

$$\rho_{\rm DM} \equiv \frac{\sum_{i=x,y,z}[2\lambda\Phi_{,i}]_{,i}}{8\pi G}, \tag{24}$$

where we use the notation $F_{,i} \equiv \partial_i F$, and the dummy index implies covariant or contra-variant derivatives with respect to x, y, z. We use the approximation that the DE part $8\pi G b\rho_{\rm DE} = \frac{\dot\varphi^2}{2N^2} + \Lambda_0 F$ is a negligible source compared to $8\pi G\rho$ from the baryons, and that

$$-\Psi = \Phi \tag{25}$$

from the spatial cross term of the Einstein equation. The above result is essentially a Poisson equation where the vector field creates an effective DM-like source term $\rho_{\rm DM}$. Rearrange the terms; the same equation becomes the MOND Poisson equation:

$$\nabla\cdot[(1-\lambda)\nabla\Phi] = 4\pi G\rho, \quad \lambda \equiv \varphi^2. \tag{26}$$

To see that $1-\lambda$ can be identified with the MOND μ_M function, first we define a value of the scalar field φ_M such that

$$F'|_{\lambda=\varphi_M^2} \equiv \frac{|\nabla\Phi|^2}{\Lambda_0}. \tag{27}$$

We find that the scalar field EOM is given as (see Equation (C2))

$$-c_\varphi^2\nabla^2\varphi = -[\Lambda_0 F' - |\nabla\Phi|^2](2N^2\varphi), \tag{28}$$

where we neglect all time-dependent terms. This equation is similar to the equation of the Yukawa potential with a screening length of c_φ/ω, $(c_\varphi^2\nabla^2 - \omega^2)\varphi = 0$. In the simplest case, we adopt $c_\varphi^2 \to 0$ to kill the Laplacian term $\nabla^2 \equiv \sum_{i=x,y,z}\partial_i\partial_i$, and we find that the equation for the scalar field becomes

$$k^2\Lambda_0\lambda^{\frac{3}{k}-1}(\lambda^{-1/k}-1)^2 \approx |\nabla\Phi|^2 \tag{29}$$

for our choice of $F(\lambda)$, where we neglect the term $\epsilon^4\varphi^2$ for a lighter expression. For $k = 3$, the equation can then be easily solved as[9]

$$\lambda \equiv \varphi^2 \to \varphi_M^2 \approx \left(1+\frac{x}{3}\right)^{-3}\Bigg|_{x=\frac{|\nabla\Phi|}{\sqrt{\Lambda_0}}}. \tag{30}$$

[9] The model with $k = 2$ gives a MOND $\mu = 1-\lambda$ $= \left(\sqrt{1+\frac{x^2}{16}}+\frac{x}{4}\right)^{-4}\Bigg|_{x=\frac{|\nabla\Phi|}{\sqrt{\Lambda_0}}}$.

To see if we recover the properties of MOND function μ_M (see $1-\varphi_M^2$ vs. x shown in Figure 1), we rewrite the solution of the scalar field as

$$1-\varphi_M^2 \equiv \mu_M \approx \begin{cases} x, & \text{where } x \equiv \frac{|\nabla\Phi|}{\sqrt{\Lambda_0}} \ll 1 \\ 1-\epsilon^2-\left(\frac{x}{3}\right)^{-3}, & \text{where } x \gg 1. \end{cases} \tag{31}$$

This is exactly the physics of MOND, if

$$\sqrt{\Lambda_0} \to a_0 \tag{32}$$

is identified with the MOND acceleration scale a_0. In the solar system or strong gravity regime, the modification factor $1-(x/3)^{-3} \sim 1$ to the Newtonian Poisson equation is small and reduces sharply. In weak gravity, applying the spherical approximation around a dwarf galaxy of mass m_b, we have $|\nabla\Phi|^2/a_0 = Gm_b r^{-2}$, and the rotation curve $V_{\rm cir}^2(r) = r\nabla\Phi$. The big success of MOND in dwarf spiral galaxies is to explain their Tully–Fisher relation $V_{\rm cir}^4(r)/(Gm_b) = a_0 \sim 10^{-10}$ m s^{-2} if $\Lambda_0 \sim (1\times10^{-10}$ m s$^{-2})^2$, which is of the order of magnitude of the observed amplitude of "the cosmological constant" effect. In the intermediate regime, our μ_M resembles the "standard" $\mu = \frac{x}{\sqrt{1+x^2}}$ function of MOND, so it will fit rotation curves of galaxies very well.

5. TEMPORAL AND SPATIAL CORRECTIONS TO MOND: OSCILLATIONS AND DIFFUSIONS

When considering merging systems like galaxy clusters, time-dependent terms $\ddot\lambda \sim \dot\lambda^2 \sim O(\omega^2)$ are important, where $\omega = O(|\mathbf{k}|\sigma)$ is the inverse of the timescale to cross a system of size $|\mathbf{k}|^{-1}$ by stars of velocity dispersion σ in unit of the speed of light. There can also be diffusion on small scale due to a pressure-like term $\nabla^2\lambda = -|\mathbf{k}|^2\lambda$.

The scalar field EOM becomes (see Equation (C2))

$$\left[\partial_t^2 - c_\varphi^2\nabla^2 + \left(1-c_\varphi^2\right)\eta\partial_t\right]\varphi = -(\varphi-\varphi_M)\,\nu^2, \tag{33}$$

where $\nu^2 \equiv 2N^2\varphi\Lambda_0\frac{F'-F'_M}{\varphi-\varphi_M} \sim 4N^2\Lambda_0F''\varphi_M^2$, and $F = F(\varphi^2)$, $F'_M \equiv F'(\varphi_M^2) = (\nabla\Phi)^2/\Lambda_0$ by definition of φ_M. We also introduce η^{-1} as a damping timescale due to coupling of φ with the Æ vector field (cf. Equation (C5) in the Appendix for the EOM of the vector field). The diffusion term $-c_\varphi^2\nabla^2 = c_\varphi^2|\mathbf{k}|^2$ is non-zero unless $c_\varphi^2 = 0$. The scalar field φ then follows the equation of a damped harmonic oscillator with a damping rate $(1-c_\varphi^2)\eta$, a slightly nonlinear restoring force $\sim -\nu^2\varphi$, and an external force $\sim \nu^2\varphi_M \sim |\nabla\Phi|^2(2N^2\varphi_M)$. Assuming that the correction due to Hubble expansion $O(3H^2+2H\eta)b'$ is negligible for a very small b', the scalar field φ eventually approaches the MOND-like static solution φ_M, thanks to the damping term with a timescale η^{-1}, which kills any history dependence. Rapid oscillations will likely keep the fluid's time-averaged property close to MOND-like solution as well.

We estimate the oscillation timescale

$$\sqrt{\frac{\varphi}{\ddot\varphi}} \sim \nu^{-1} = (2N)^{-1}(\Lambda_0F''\lambda)^{-1/2}|_{\lambda=\varphi_M^2} \sim \frac{10^8}{N}\times 300\text{ yr}, \tag{34}$$

which is about 10^9 yr if $N \sim 10$. Here we assume $\varphi_M = \sqrt{\lambda} = O(1) = F''$ for systems of mild gravity ($\sim10^{-8}$ cm s^{-2}, e.g., clusters; for systems of stronger gravity, the timescale is perhaps longer). In the process of damping, there will be a correction to the MOND μ_M function by the q term, heuristically, $1-\varphi^2 = 1-\varphi_M^2$, if

$$x \to \sqrt{\frac{|\nabla\Phi|^2}{\Lambda_0} + \frac{q}{2N^2\Lambda_0}}, \tag{35}$$

where $q \equiv \left[\partial_t^2 - c_\varphi^2\nabla^2 + (1-c_\varphi^2)\eta\partial_t\right]$ is an operator. In tidally acting systems, the value for φ will oscillate between its pre-merging value and its equilibrium value.

Models with a small N would not give MOND, e.g., if $N = 1$–10, $\pi\nu^{-1} \sim (100-10)$ Gyr, then the universe would be too young dynamically to have a precise MOND effect in galaxies because φ would not have enough time to respond to the formation of galaxies. Rather φ would lag behind and might remain close to its cosmological average,

$$\varphi \sim \bar\varphi, \tag{36}$$

which would mean a boost of the gravity of the baryon by a constant factor $(1-\bar\varphi^2)^{-1}$ everywhere.

6. GENERIC PROPERTIES OF DARK FLUID

It is still uncertain whether the time-dependent correction and a possible diffusion term are enough to help MOND to explain the Bullet clusters (Angus et al. 2007; Angus & McGaugh 2008). However, it seems robust that the *dark fluid*—described by the field $Æ^a\varphi(\lambda)$—is generally out of phase from the baryonic fluid. The dynamics of galaxies inside the dark fluid are governed by five variables, two for the scalar field $\varphi = \sqrt{\lambda}$, the metric or potential $\Phi = -\Psi = \frac{g_{00}-1}{2} = Æ_0 - 1$, and three for the vector field perturbation part $Æ_i = Æ_i - u_i$.[10] These evolve according to the five coupled Equations (C2), (C5), and (C9) in the absence of the Hubble expansion. The proof is given in Appendix C; the following is a summary of the equations:

$$\left[\frac{\partial_t^2 - c_\varphi^2\partial_i^2 - \left(1-c_\varphi^2\right)(\partial_i æ_i)\partial_t}{2N^2\Lambda_0}\right]\varphi - \frac{(\partial_t æ_i - \partial_i\Phi)^2}{\Lambda_0}\varphi + F'\varphi = 0, \tag{37}$$

$$\partial_t[\lambda(\partial_t æ_i - \partial_i\Phi)] + \partial_i\left[\frac{(\partial_i æ_i + 3\partial_t\Phi)(-c_2)}{2}\right] + \frac{1-c_\varphi^2}{2N^2}\partial_t\varphi\partial_i\varphi = 0, \tag{38}$$

$$\partial_i[\lambda\partial_t æ_i + (1-\lambda)\partial_i\Phi] = +4\pi G\rho^{\rm bary}, \tag{39}$$

where the time-derivative terms are all moved to the left-hand side, the summation over dummy index $i = x, y, z$ is implicit, $\sqrt{\Lambda_0} \sim 10^{-10}$ms^{-2}, $c_\varphi \sim 1$, and N are constants, and we used $\varphi = \sqrt{\lambda}$, $\eta = -\partial_i æ_i$, $c_2 = -\frac{2}{3\lambda}$, and $c_4 = 2\lambda$.

These equations, when supplemented by the continuity equation and the three momentum equations for the normal baryonic

[10] It is not necessary to compute $Æ_a = u_a + æ_a$ (cf. Equation (B13)), or to evolve the auxiliary unit vector field $u_a = (1+\Phi, u_1, u_2, u_3)$ with the geodesics of test particles $u^a\nabla_a u_i = A_i = -\partial_i\Phi$ explicitly. All 3+1 formulated physical quantities are expressed in terms of $æ_a = (0, æ_1, æ_2, æ_3)$, and the local volume expansion rate $\theta = \nabla_a u^a = 3H + 3\dot\Psi$ and the local acceleration A_i.

fluid described by $\rho^{\rm bary}$ and v_i with a certain equation of state $P^{\rm bary}(\rho^{\rm bary})$, completely specify the dynamics.

The dark fluid has two types of deviations from MOND in general.

1. The dark fluid has a generic scalar field vector oscillation and damping timescales $\sim O(\nu^{-1}) \sim O(\eta^{-1}) \sim O(N)$ times the orbital crossing time roughly unless the system is hotter than $c/\sqrt{N}$ or the forces are in resonance. A very fast damping would mean an almost instantaneous relation between gravity and the scalar field $1-\varphi^2$, as the μ_M in the classical Bekenstein & Milgrom (1984) modified gravity interpretation of MOND. A slow damping would mean a history-dependent relation, reminiscent of Milgrom's modified inertia interpretation of MOND: the dark fluid adds a dynamically varying inertia around the baryons that it surrounds. A possible test could be in galaxies with rotating bar(s), where there could be a phase lag between the bar and the effective DM (Debattista & Sellwood 1998). This has intriguing consequences to the bar's pattern speed because of nontrivial corrections to the MOND pictures of dynamical friction (Ciotti & Binney 2004; Nipoti et al. 2008; Tiret & Combes 2008); the properties of the dark fluid are in between that of real particle dark halo and that naively expected from MOND.
2. The dark fluid has a pressure, controlled by a propagation speed c_φ, where the speed of light is unity here, and the dark fluid can be made cold by $c_\varphi^2 \sim 0$, or hot by $c_\varphi^2 \sim 1$, or superluminal by $c_\varphi^2 \geqslant 1$. The φ would no longer be a function of the local gravity at $\mathbf{r}$ (as in MOND), rather it is a weighted average of a volume of all points $\mathbf{r}_1$ by a Yukawa-type screening function $\exp(-\frac{\nu|\mathbf{r}_1-\mathbf{r}|}{c_\varphi})$, where

$$\text{Screening length} = c_\varphi \nu^{-1} \sim c_\varphi \frac{10^8}{N} \times 300 \text{ lt yr}. \quad (40)$$

Note this spatial correction to MOND can exist even in static systems; even a small pressure term with $c_\varphi^2 \neq 0$ might smooth out MOND effects on small-scale structures (wide binaries, star clusters, dwarf galaxies), where the wave number $|\mathbf{k}|^2$ is much bigger than in galaxy clusters. The screening length can be set at ~ 100 pc for either an $N \sim 10^8$, $c_\varphi \sim 3\times 10^5$ km s^{-1} hot dark fluid or an $N \sim 10^4$ and $c_\varphi \sim 30$ km s^{-1} cold dark fluid. This scale, 100 pc, is a scale dividing dense star clusters and fluffy dwarf galaxies. Observational DM effects are only seen in the universe on scales larger than 100 pc. It has been challenging for MOND to explain this observed scale (Zhao 2005; Sanchez-Salcedo & Hernandez 2007; Baumgardt et al. 2005).

In conclusion, we find a framework of dark fluid theories where MOND corresponds to a special choice of potentials or mass for the vector field. The dark fluid can run cold or hot depending on the sound speed c_φ (which could even be a running function of the vector field). These theories degenerate into scalar field theories for DE effects in the Hubble expansion. It is possible to create an exact $w = -1$ DE effect. The scale $a_0 = \sqrt{\Lambda_0}$ in MOND in equilibrium spiral galaxies derives its physics from the amplitude of the DE Λ_0. MOND or DM effects are hence indications of a non-uniform DE fluid described generally by a vector field $\varphi Æ^a$. For non-equilibrium systems like the Bullet clusters or galaxies with satellites, the properties of the dark fluid do not follow exactly the usual expectations of MOND or cold/hot DM, but (not so surprisingly) in between.

H.S.Z. acknowledges Martin Feix and Gil Esposito-Farse for discussions on PPN and causality issues.

APPENDIX A

MOTIVATING A GENERIC LAGRANGIAN RELATING VARIOUS THEORIES WITH ONE FIELD

We first motivate a generic Lagrangian behind our specific Lagrangian $L(\varphi, Æ^a)$ for the dark fluid. The idea is mainly to relate various theories together with only one field. We start from any vector field of 4 degrees of freedom, denoted by Z^a. It has generally a variable or dynamic norm

$$\lambda \equiv \varphi^2 \equiv |g_{ab} Z^a Z^b|, \quad (\mathrm{A}1)$$

where we adopt the convention of Einstein summation of identical upper and lower indices. The field λ is an auxiliary scalar field characterizing the norm of the vector field Z^a, hence is not an independent dynamical freedom (we will return to this point below). A generic coupling of the vector field Z^a with the spacetime is through the contractions among the $Z^a Z^b$ tensor, the g_{ab} metric tensor, and the Ricci tensor R_{ab}. Hence the most generic theory of the vector should have an action $S = -(16\pi G)^{-1} \int dx^4 \sqrt{-g}(\mathcal{L}_m + \mathcal{L})$, where the Lagrangian density containing a term $\mathcal{L}_m$ due to matter, and the term

$$\mathcal{L} = 2\Lambda_0 F(\lambda) + [g^{ab} f_1 - Z^a Z^b \varphi^{-2} f_2] R_{ab} + f_3 \nabla^a \lambda \nabla_a \lambda + \cdots, \quad (\mathrm{A}2)$$

where $F(\lambda)$ and f_i are dimensionless functions of λ, and the typical energy density set by a constant Λ_0, which is the only dimensional scale in the dark fluid. Note that $R = g^{ab} R_{ab}$ is the Ricci scalar, and has the dimension of Λ_0.

The above Lagrangian is generic enough, and many DE models can be derived from it. For example, the terms $\Lambda_0(\lambda-1)^{1/(1+n)} + f_1 R$ with $f_1 = \lambda$ can lead to an $R + \mathrm{const}/R^n$ (Li et al. 2008) gravity (as could be checked by solving λ from the EOM of λ and then substituting back into the original Lagrangian); and the terms $f_1 R + f_3 \nabla^a \lambda \nabla_a \lambda$ with $f_1 = \lambda$ and $f_3 = -cst\lambda^{-1}$ would lead to the Brans–Dicke theory of gravity. The usual quintessence theories can be recovered from the terms $\Lambda_0 F(\lambda) + R + \nabla^a \lambda \nabla_a \lambda$ with a suitable potential $F(\lambda)$.

The essential dynamics of a cosmological vector field is described by the term $f_2 \varphi^{-2} Z^a Z^b R_{ab}$, which can be moulded into very different but equivalent forms, e.g., $\mathcal{L} = K^{ac}_{bd} \nabla_a Z^b \nabla_c Z^d + f_1 R + 2\Lambda_0 F(\varphi^2) + 4 f_3 \varphi^2 \nabla^a \varphi \nabla_a \varphi + \cdots$, where[11] the tensor K^{ac}_{bd} is a function of φ (Ferreira et al. 2007; Halle et al. 2008). If we require that the field Z^a has a unit norm guaranteed by a Lagrange multiplier, we would recover Einstein–Aether theory (Jacobson & Mattingly 2001) and its generalizations (Zlosnik et al. 2007; Li et al. 2008).

On the other hand, the generality of the Lagrangian can also create problems, especially a coupling like $Z^a Z^b R_{ab} \varphi^{-2} f_2$, which can dangerously widen the cone of causality. Care must be taken in selecting the coefficients f_1, f_2, f_3, etc. as the package is simply too broad to be healthy as a whole. It is necessary to eliminate, by hand, certain mathematically allowed terms to pick a safe subset of the theory. To facilitate the selection, let us decompose the four dynamical degrees of freedom in the vector

[11] Note that $Z^b R_{ab} = (-\nabla_a \nabla_c Z^c + \nabla_c \nabla_a Z^c)$ by definition of the Ricci tensor, so the term $\int dx^4 \sqrt{-g} Z^a Z^b R_{ab} \varphi^{-2} f_2(\varphi) \rightarrow \int dx^4 \sqrt{-g} K^{ac}_{bd} \nabla_a Z^b \nabla_c Z^d$, where a full divergence term is dropped after turning into a surface integral.

field Z^a into a scalar degree and three dynamical degrees of a unit-norm vector field $Æ^a$,

$$Æ^a \equiv \frac{Z^a}{\varphi}. \tag{A3}$$

These three degrees of freedom for the unit-norm vector field $Æ^a$ could be identified by a similar analysis as that of Graesser et al. (2005). This decomposition makes sense unless the norm of Z^a is zero, i.e., $\varphi^2 = 0$, at which point the choice of $Æ^a$ is not unique. If $\varphi^2 = |Z^a Z_a|$ is always positive, then the vector field $Æ^a$ will be space-like/time-like if Z^a is space-like/time-like. A space-like vector field picks out a preferred spatial direction, which could violate various constraints, so it is safer to limit our Z^a (hence $Æ^a$) to be time-like. Because we do not desire Z^a to change from time-like to light-like ($\varphi^2 = 0$) in the solar system, care must be taken to the choice of the potential F, f_3, etc. to prevent φ from reaching zero (see Appendix D about the function F).

Finally, we add a cautionary note:

1. Any Lagrangian of the form $\mathcal{L}(Z^a)$ could be expressed as $\mathcal{L}(Æ^a, \varphi)$ and vice versa; one only needs to substitute $\varphi = \sqrt{|Z^b Z_b|}$ and $Æ^a = Z^a/\sqrt{|Z^b Z_b|}$. The expressions can be very compact in one and very lengthy in another. It seems safe to speak of Z^a as a unifying field at least in the effective sense if not fundamentally meaningful.
2. A vector field Z^a of a dynamical norm would be time-like if it is actually related to a unit-norm time-like field $Æ^a$ by a metric redefinition. The redefined metric can absorb the norm (see Section 4.2 of Zhao 2008b); hence, what appears as a unit vector in one metric can have a dynamical norm in another metric. For the same reason, Zlosnik et al. (2007) found a lengthy Lagrangian when combining the unit vector and scalar of TeVeS into one time-like non-unit vector. Another possibility is for Z^a to be a complex unit vector related to neutrino mass (cf. Zhao 2008b).

A detailed discussion on various interpretations of Z^a is beyond this paper. While it is conceptually satisfactory to relate all dynamics of a Lagrangian to one vector field Z^a, a sensible Lagrangian is more readily constructed in a compact way using the $Æ^a$ unit vector field of 3 degrees of freedom and the φ field of 1 degree of freedom instead of Z^a of 4 degrees of freedom. We shall restrict to the $Æ^a$ and φ notations only in the main text of the paper.

APPENDIX B

EOM AND STRESS TENSOR FOR THE GENERAL LAGRANGIAN

The most general Lagrangian that we consider (Equation (1)) can be cast to the form

$$\begin{aligned}\mathcal{L} = {} & (1 + e_0\varphi^2)R + V(\varphi^2) + (Æ^a Æ_a - 1)L^* + \sum_{j=1,4} c_j K_j \\ & + [-d_1 g^{ab} + d_2(g^{ab} - Æ^a Æ^b)]\nabla_a\varphi\nabla_b\varphi \\ & - (d_3 Æ_a g_{bc} + d_4 Æ_b g_{ac})\nabla^b\varphi\nabla^a Æ^c,\end{aligned} \tag{B1}$$

where R is the Ricci scalar, and L^* is the Lagrange multiplier (a kind of potential), where

$$K_1 = \nabla_a Æ_c \nabla^a Æ^c \tag{B2}$$

$$K_2 = (\nabla_a Æ^a)^2 \tag{B3}$$

$$K_3 = \nabla_c Æ_a \tag{B4}$$

$$K_4 = \nabla_\parallel Æ_c \nabla^a Æ^c, \tag{B5}$$

where $\nabla^a = g^{ac}\nabla_c$, $Æ_a = g_{ac}Æ^c$. We use the following shorthand notations $d_1 = 1/N^2$, $d_2 = (1 - c_\varphi^2)/N^2$, $V = 2\lambda_0 F$, $d_3 = e_1 d_4 = e_2$, $c_4 = 2\varphi^2$, $c_{14} = c_1 + c_4$, $d_{34} = d_3 + d_4$, and $\alpha = c_1 + 3c_2 + c_3$, $\dot\varphi = \partial_t\varphi$. The term $e_0\varphi^2 R$ in the Lagrangian density is the scalar–tensor term which we do not use in this work, and the terms in field equations due to this can be found easily in the literature, so here we neglect it by setting $e_0 = 0$.

The stress–energy tensor is

$$8\pi G T_{ab} = 8\pi G T_{ab}^{\varphi} + 8\pi G T_{ab}^{Æ}, \tag{B6}$$

where

$$\begin{aligned}8\pi G T_{ab}^{\varphi} = {} & g_{ab}V + (d_1 - d_2)\left[\nabla_a\varphi\nabla_b\varphi - \frac{1}{2}g_{ab}\nabla_c\varphi\nabla^c\varphi\right] \\ & + \nabla_c\left[d_3 Æ^c(Æ_a\nabla_b\varphi + Æ_b\nabla_a\varphi) - \frac{d_3}{2}Æ_a Æ_b \nabla^c\varphi\right] \\ & + \frac{d_4 - d_3}{2}(\nabla_d\varphi)[(Æ^c\nabla_c Æ^d)(g_{ab} - Æ_a Æ_b) \\ & + Æ^d(\nabla_c Æ^c)Æ_a Æ_b] \\ & + Æ^c Æ^d\left[\frac{d_2 - d_3'}{2}\nabla_c\varphi\nabla_d\varphi - \frac{d_3}{2}\nabla_c\nabla_d\varphi\right]Æ_a Æ_b \\ & + Æ^c Æ^d\left[\frac{-d_2 + d_4'}{2}\nabla_c\varphi\nabla_d\varphi + \frac{d_4}{2}\nabla_c\nabla_d\varphi\right] \\ & \times (g_{ab} - Æ_a Æ_b),\end{aligned} \tag{B7}$$

$$\begin{aligned}8\pi G T_{ab}^{Æ} = {} & \frac{1}{2}g_{ab}\mathcal{K} + [Æ_d\nabla_c J^{cd}]Æ_a Æ_b \\ & + \nabla_c\left[Æ_{(a}J^{c}_{\ b)} - Æ^c J_{(ab)} - Æ_{(a}J_{\ b)}^{c}\right] \\ & + c_1[\nabla^c Æ_a \nabla_c Æ_b - \nabla_a Æ^c \nabla_b Æ_c] \\ & + c_4 Æ^c Æ^d[\nabla_c Æ_a \nabla_d Æ_b - Æ_a Æ_b \nabla_c Æ_e \nabla_d Æ^e],\end{aligned}$$

where we have defined

$$J^a_{\ c} \equiv K^{ab}_{\ \ cd}\nabla_b Æ^d, \quad \mathcal{K} \equiv K^{ab}_{\ \ cd}\nabla_a Æ^c \nabla_b Æ^d = \sum_j c_j K_j \tag{B9}$$

$$K^{ab}_{\ \ cd} = c_1 g^{ab} g_{cd} + c_2\delta^a_c\delta^b_d + c_3\delta^a_d\delta^b_c + c_4 Æ^a Æ^b g_{cd}. \tag{B10}$$

The scalar EOM is

$$\begin{aligned}-\left[V' + \sum_j \frac{c_j'}{2}K_j\right] = {} & g^{ab}\left[\frac{d_1'}{2}\nabla_a\varphi\nabla_b\varphi + d_1\nabla_a\nabla_b\varphi\right] \\ & + \frac{d_3}{2}\nabla_b(Æ^a\nabla_a Æ^b) + \frac{d_4}{2}\nabla_a(Æ^a\nabla_b Æ^b) - (g^{ab} - Æ^a Æ^b) \\ & \times\left[\frac{d_2'}{2}\nabla_a\varphi\nabla_b\varphi + d_2\nabla_a\nabla_b\varphi\right] + d_2\nabla_a\nabla_b(Æ^a Æ^b).\end{aligned} \tag{B11}$$

The vector EOM is

$$\begin{aligned}-L^* Æ_a = {} & -\nabla_b J^b_{\ a} + c_4\nabla_a Æ_c(Æ^b\nabla_b Æ^c) \\ & + \left[\frac{d_3' + d_4'}{2} - d_2\right](Æ^b\nabla_b\varphi)\nabla_a\varphi \\ & + \frac{d_3 - d_4}{2}[(\nabla_b Æ^b)\nabla_a\varphi - (\nabla_a Æ^b)\nabla_b\varphi] \\ & + \frac{d_3 + d_4}{2}Æ^b\nabla_b\nabla_a\varphi,\end{aligned} \tag{B12}$$

where the dependence on the Lagrangian multiplier L^* can be eliminated by multiplying both sides by $(g^{ab} - Æ^a Æ^b)$ and contract over index a. Alternatively, L^* can be computed by multiplying both sides by $Æ^a$ and contract over the index a.

In the following, we list the field equations in the weak field limit, applicable to galaxies in an expanding background metric. In galaxies, the metric is nearly flat but the overdensity of matter can be much bigger than unity. We use the 3+1 formulation (see Appendix E). We write

$$Æ_a = u_a + æ_a, \tag{B13}$$

where u_a is the four-velocity of an observer on geodesic, and like $Æ_a$ is a unit-norm time-like vector field. Since the background universe is observed to be homogeneous and isotropic to high precision, the difference vector field, $æ_a$, between them should be of first order in perturbation. The relations

$$Æ_a Æ^a = u_a u^a = 1 \rightarrow u^a æ_a = 0 \tag{B14}$$

which means that the field $æ_a$ is perpendicular to the four-velocity of the observer (see Li et al. 2008).

As discussed in Halle et al. (2008), it is highly nontrivial to keep track of the necessary orders of the equation to be valid both for static highly nonlinear galaxies in a static empty universe and for linear perturbations around a uniformly expanding universe. Here we treat the zeroth order being a static galaxy in a nonlinear overdensity. Any temporal changing terms or terms involving Hubble expansion are at least first-order perturbations. Besides the first order, we have kept some second-order terms whenever the first order can become zero in a static galaxy. For a lighter notation, we also define intermediate variables $\eta \equiv \hat{\nabla}^a æ_a$, $B_a \equiv A_a + \dot{æ}_a + \frac{\theta}{3} æ_a$, $Q_{ab} \equiv c_{13}(\sigma_{ab} + \hat{\nabla}_{\langle a} æ_{b\rangle})$.

The scalar EOM is

$$\begin{aligned}\left[d_1 \partial_t^2 + (d_1 - d_2)\hat{\nabla}^2\right]\varphi + (d_1\theta + d_2\eta)\,\dot{\varphi} = -\frac{d_3}{2}\hat{\nabla}^a B_a \\ - \frac{d_4}{2}[(\dot{\theta} + \dot{\eta}) + (\theta + \eta)^2] - \frac{\partial}{\partial\varphi}\left[V + \sum c_i K_i\right],\end{aligned} \tag{B15}$$

where $\sum c_i K_i = -\frac{c_{14}}{2} B^a B_a + \frac{\alpha}{6}(H + \eta)^2 + \cdots$, which includes second-order terms, non-negligible for a nearly static galaxy. We will not go into complete expressions for the second-order terms, which are beyond this paper but have been discussed in Halle et al. (2008).

The vector EOM is

$$\begin{aligned}\left(\partial_t + \frac{2\theta}{3}\right)(c_{14}B_a) + \hat{\nabla}_a\left[\frac{\alpha(\theta + \eta)}{3}\right] - æ_a \partial_t\left(\frac{\alpha\theta}{3}\right) = -\hat{\nabla}^b Q_{ab} \\ + \partial_t\left[\frac{d_3 + d_4}{2}(\hat{\nabla}_a\varphi - \dot{\varphi}æ_a)\right] - \left[d_2\dot{\varphi} - \frac{d_3\theta}{6} + \frac{d_4\theta}{2}\right] \\ \times (\hat{\nabla}_a\varphi - \dot{\varphi}æ_a).\end{aligned} \tag{B16}$$

The stress–tensor components due to the Æ and φ fields are given by

$$8\pi G\pi_{ab}^{\varphi Æ} = -\dot{Q}_{ab} - \theta Q_{ab}, \tag{B17}$$

$$\begin{aligned}8\pi G q_a^{\varphi Æ} = d_1\dot{\varphi}\hat{\nabla}_a\varphi + \hat{\nabla}_a\frac{\alpha(\theta + \eta)}{3} \\ + (\theta - \partial_t)\left(\frac{d_4}{2}\hat{\nabla}_a\varphi\right) + \hat{\nabla}^b Q_{ab},\end{aligned} \tag{B18}$$

$$\begin{aligned}8\pi G p^{\varphi Æ} = \left[\frac{d_1}{2}\dot{\varphi}^2 - V + \frac{\alpha}{6}(\theta + \eta)^2\right] \\ + \partial_t\left[\frac{\alpha(\theta + \eta)}{3}\right] - \partial_t\left(\frac{d_4}{2}\dot{\varphi}\right),\end{aligned} \tag{B19}$$

$$\begin{aligned}8\pi G\rho^{\varphi Æ} = \hat{\nabla}^a(c_{14}B_a) + \left[\frac{d_1}{2}\dot{\varphi}^2 + V - \frac{\alpha}{6}(\theta + \eta)^2\right] \\ + \frac{d_3}{2}\dot{\varphi}\eta - \hat{\nabla}^a\left(\frac{d_3}{2}\hat{\nabla}_a\varphi\right) + \frac{d_4}{2}\dot{\varphi}(\theta + \eta).\end{aligned} \tag{B20}$$

One can verify that these stress-tensor components satisfy the conservation equations, $\dot{\rho} + (\rho + p)\theta + \hat{\nabla}^a q_a = 0$ and $\dot{q}_a + \frac{4}{3}\theta q_a + (\rho + p)A_a - \hat{\nabla}_a p + \hat{\nabla}^b\pi_{ab} = 0$, where the superscript $\varphi Æ$ is omitted.

The $\nabla\varphi\nabla Æ$ coupling leads to terms like $O(d_3\nabla^2\varphi)$ in the density $\rho^{\varphi Æ}$, which do not appear to have a lower bound, hence could lead to unbound Hamiltonian if d_3 and d_4 were not set to zero. Also a non-zero c_{13} term clearly creates anisotropic stress, and hence excites the spin-1 mode. The $\varphi^2 R$ term could lead to a Hamiltonian density $O(\nabla^2\varphi)$, which can be unbound when φ is big. Undesirable modified gravity effects in the solar system could also be created by the term $Æ^a\nabla_a Æ^b\nabla_b\varphi \sim O(1)\partial_i g_{00}\partial_i\varphi$, where ∂_i stands for spatial derivatives and can be very big on small scales. The term $Æ^b\nabla_a Æ^a\nabla_b\varphi \sim O(H)\partial_t\varphi$ can have big influence for structure formation with little effects on the solar system, and hence might be desirable. For above considerations, we set $d_3 = d_4 = c_1 = c_3 = 0 = f$, in which case the equations are greatly simplified.

APPENDIX C

EQUATIONS AND DAMPING RATES OF SCALAR–VECTOR FIELDS IN GALAXIES

Here we give some relations between our 3 + 1 variables and the conventional gravitational potentials in the Newtonian gauge:

$$ds^2 = (1 + 2\Phi)dt^2 - a^2(t)(1 + 2\Psi)(dx^2 + dy^2 + dz^2),\ a(t) = 1. \tag{C1}$$

In the following, we shall neglect the Hubble expansion, and limit the discussion to our choice of parameters with $d_3 = d_4 = 0$, and $c_{13} = \alpha - 3c_2 = 0$. The metric has neither spin-2 nor spin-1 mode, and has purely spin-0 mode due to variations of Newtonian potentials $\Phi(t, x, y, z) = -\Psi(t, x, y, z)$. We neglect any vorticity generated by the scalar field and non-relativistic baryons.

The scalar EOM reduces to Equation (33):

$$\left[\partial_t^2 - c_\varphi^2\nabla^2 + \left(1 - c_\varphi^2\right)\eta\partial_t\right]\varphi = [-\Lambda_0 F' + (\nabla\Phi - \dot{æ})^2](2N^2\varphi), \tag{C2}$$

which has the form of a damped harmonic oscillator with the damping rate η determined by

$$\eta \equiv \hat{\nabla}^j æ_j = -\partial_j æ_j = -\partial_j\partial_j Y, \tag{C3}$$

with an implicit summation of the dummy index $j = x, y, z$. In the following, we consider mainly the evolution of the vector field perturbation $æ_j = \partial_j Y \ll 1$ in the compressional spin-0 mode with a potential Y.

That $c_{13} = 0$ means that two Newtonian potentials are related by $\Psi = -\Phi$ and in the vector field EOM and heat flux we can drop terms such as Q_{ab}. The terms of σ_{ab}, ϖ_{ab} in the constraint Equation (E5) can be dropped if we assume negligible (non-relativistic) vorticity and shear due to the baryon heat flux, so

$$-2\dot{A}_i = 2\partial_i\dot{\Phi} = -2\partial_i\dot{\Psi} = -\partial_i\left(\frac{2\theta}{3}\right)$$
$$= 8\pi G q_i^{\text{bary}} + \frac{1}{N^2}\dot{\varphi}\partial_i\varphi + \partial_i\left[c_2(\eta+\theta)\right] \quad \text{(C4)}$$

gives the relation of the variables $\theta = \nabla_a u^a$ and $A_i = u^a\nabla_a u_i$ in the 3+1 formalism with the gravitational potentials in the conformal Newtonian gauge; for the last equation, Equation (E5) and Equation (B18) are combined, with the total heat flux $q_i = q_i^{\varphi Æ} + q_i^{\text{bary}}$, where the baryon heat flux contribution $q_i^{\text{bary}} = \rho^{\text{bary}} v_i$ for a baryon density ρ^{bary} and velocity v_i.

Dropping higher order terms, the vector field EOM (Equation (B16)) becomes

$$\partial_t\left[2\lambda B_i\right] = -\partial_i\left[c_2(\eta+\theta)\right] - \frac{1-c_\varphi^2}{N^2}\dot{\varphi}\partial_i\varphi, \quad \text{(C5)}$$

for the index $i = x, y, z$, where $B_i = -\partial_i\Phi + \dot{æ}_i$, and we approximated $\frac{[\varphi_{,i} - \dot{\varphi}æ_i]}{N^2} \sim \frac{\partial_i\varphi}{N^2}$ because $|æ_i| \ll 1$, and $|\dot{\varphi}| \ll |\varphi_{,i}|$ inside the causal horizon. Using the expression for $8\pi G q_i^{\varphi Æ}$ in Equation (C4), the EOM becomes

$$[2\lambda(\ddot{æ}_i - \partial_i\dot{\Phi}) + 2\dot{\lambda}\dot{æ}_i] = -[2\partial_i\dot{\Phi} - 8\pi G\rho^{\text{bary}}v_i] + \frac{c_\varphi^2}{N^2}\dot{\varphi}\partial_i\varphi. \quad \text{(C6)}$$

The equation can be simplified as

$$2[\lambda\ddot{æ}_i + \dot{\lambda}\dot{æ}_i - \dot{\lambda}\partial_i\Phi + (1-\lambda)\partial_i\dot{\Phi}] = S_i,$$
$$S_i \equiv \left(\frac{c_\varphi^2\partial_i\lambda}{4N^2\lambda}\right)\dot{\lambda} + 8\pi G\rho^{\text{bary}}v_i. \quad \text{(C7)}$$

Thus, the vector perturbation $æ_i$ and the scalar field $\phi = \sqrt{\lambda}$ evolve as two coupled damped harmonic oscillators with damping rates $\frac{\dot{\lambda}}{\lambda}$ and η, respectively. The source term on the right-hand side of the EOM of the vector would be zero, $S_i = 0$, for a very cold dark fluid $c_\varphi^2/N^2 = 0$ and in the vacuum where $\rho^{\text{bary}} = 0$.

Finally, in the absence of the Hubble expansion, the Poisson equation becomes

$$-2\partial_i^2\Psi = 8\pi G\rho^{\text{bary}} + \left[\hat{\nabla}^i(2\lambda B_i) + \frac{1}{2N^2}\dot{\varphi}^2 + V - \frac{\alpha}{6}(\theta+\eta)^2\right], \quad \text{(C8)}$$

where $-\Psi = \Phi$, and the right-hand side terms in square brackets are from $8\pi G\rho^{\varphi Æ}$ (cf. Equation (B20)), where $\hat{\nabla}^i = -\partial_i$, $B_i = (-\partial_i\Phi + \dot{æ}_i)$, and the quadratic term $(\theta+\eta)^2 \sim [O(\dot{\Phi}) + O(N^{-2}\dot{\varphi})]^2$. For the study of the (small) galactic scale dynamics, we can safely drop all source terms involving $\dot{\varphi}^2/N^2$, $\dot{\Phi}^2$, V, which are all relativistic corrections of the order of $V \sim \Lambda_0$. Therefore, the Poisson equation in time-dependent systems can be approximated as

$$\partial_i^2\Phi = 4\pi G\rho^{\text{bary}} - \partial_i\left[\lambda(-\partial_i\Phi + \dot{æ}_i)\right]. \quad \text{(C9)}$$

This equation reduces to the form of MOND,

$$\partial_i[(1-\lambda)\partial_i\Phi] = 4\pi G\rho^{\text{bary}} + O(N^{-2}|\mathbf{k}|^2), \quad \text{(C10)}$$

in steady-state systems without the Hubble expansion, in which case we can set $\dot{æ}_i = 0$. The scalar Equation (C2) for $\lambda = \varphi^2$ recovers MOND only if $c_\varphi = \eta = 0$ and $\Lambda_0 = a^2$, in which case $(\nabla\Phi)/a_0 = \sqrt{F'(\lambda)} = 3(\lambda^{-1/3} - 1)$ for our choice $F(\lambda)$ with a negligible ϵ (Equation (6)). However, we shall show in the following the importance of corrections due to η in the scalar field equation, and $\dot{æ}_i$ in the Poisson equation.

C.1. Approximate Solutions

The evolution equations can be summarized as follows:

$$\frac{1}{(2N^2\Lambda_0\sqrt{\lambda})}\left[\partial_t^2 - c_\varphi^2\nabla^2 - (1-c_\varphi^2)(\partial_iæ_i)\partial_t\right]\sqrt{\lambda}$$
$$= -F' + \frac{(\partial_i\Phi - \partial_tæ_i)^2}{\Lambda_0}, \quad \text{(C11)}$$

$$\partial_t(\lambda\partial_tæ_i) + (1-\lambda)\partial_t\partial_i\Phi - \partial_t\lambda\partial_i\Phi = \left(\frac{c_\varphi^2\partial_i\lambda\partial_t\lambda}{8N^2\lambda}\right) + 4\pi G\rho^{\text{bary}}v_i, \quad \text{(C12)}$$

$$\partial_i\left[(1-\lambda)\partial_i\Phi\right] = -\partial_i[\lambda\partial_tæ_i] + 4\pi G\rho^{\text{bary}}. \quad \text{(C13)}$$

There is also a constraint on Equation (C4):

$$\partial_j\left[(2+3c_2)\partial_t\Phi + c_2\partial_iæ_i\right] - \frac{\partial_t\varphi\partial_j\varphi}{N^2} = +8\pi G\rho^{\text{bary}}v_j, \quad \text{(C14)}$$

which is not independent from the combination of the above equations and the continuity equation of the baryon fluid; here $c_2 = -\frac{2}{3\lambda}$ for our choice.

The above equations are hard to solve in general. Nevertheless, Equation (C4) can be heuristically integrated to a simpler relation:

$$-\dot{\Phi} = \dot{\Psi} = \frac{\theta}{3} \sim -\frac{c_2}{(2+3c_2)}\eta - \dot{\Phi}^{\varphi\,\text{bary}}, \quad \text{(C15)}$$

where $\dot{\Phi}^{\varphi\,\text{bary}}$ is a fudged contribution to $\dot{\Phi}$ due to the evolution of the scalar field and the movement of baryons. We can then express $\dot{\Phi}_{,i}$ in terms of $-\eta_{,i} = æ_{j,ji}$. and the EOM can be cast to the form

$$(2\lambda)\left[\delta_i^j\left(\partial_t^2 + \frac{\dot{\lambda}}{\lambda}\partial_t\right) - c_0^2(\partial_i\partial_j + \partial_i\xi\partial_j)\right]æ_j = \tilde{S}_i, \quad \text{(C16)}$$

where

$$c_0^2 = \frac{(1-\lambda)c_2}{(2+3c_2)\lambda}, \quad \text{(C17)}$$

$\xi \equiv \ln\frac{c_2}{2+3c_2}$, and $\tilde{S}_i = \partial_i O(\dot{\Phi}^{\varphi\,\text{bary}}) \sim O(N^{-2}\dot{\lambda}\partial_i\ln\lambda) + O(8\pi G\rho^{\text{bary}}v_i)$. Note that δ_i^j equals unity if $i = j$ or zero otherwise, and the vector field $æ_i$ appears to track the evolution of the baryons and the scalar field λ.

To see the tracking to baryons, we take can take the limit that $N \to \infty$, hence treat the scalar field as constant, so $\dot{\lambda} = 0 = \partial_i\xi$, the EOM of the vector perturbation can be approximated as

$$\left[\partial_t^2 - c_0^2\nabla^2\right]æ_i \sim \frac{(2\lambda+3c_2)8\pi G}{2\lambda(2+3c_2)}\rho^{\text{bary}}v_i, \text{ if } N \to \infty. \quad \text{(C18)}$$

We used $c_0^2\partial_i\partial_j æ_j = c_0^2\partial_i\partial_j\partial_j Y = c_0^2\partial_i\nabla^2 Y = c_0^2\nabla^2 æ_i$ since $æ_i = \partial_i Y$ for the spin-0 perturbations. Hence, we recover the propagation equation $[\partial_t^2 - c_0^2\nabla^2]æ_i = 0$ of Jacobson's Einstein–Aether theory in vacuum. Indeed, c_0^2 plays the role of sound speed squared for the spin-0 wave in vacuum. The vector field $æ_i$ tracks baryon current vector $\rho^{\rm bary} v_i$ with Equation (C18) like the magnetic vector potential tracks the moving electric charges. To estimate η, we apply $-(2\lambda)^{-1}\partial_i$ to both sides of the EOM of the vector (Equation (C16) or Equation (C18)), and sum over i, and we get an equation for η:

$$\left[\partial_t^2 - c_0^2\nabla^2\right]\eta \sim \frac{(2\lambda + 3c_2)8\pi G}{2\lambda(2+3c_2)}\partial_t\rho^{\rm bary}, \tag{C19}$$

where we used $\partial_i(4\pi G\rho^{\rm bary} v_i) = -\partial_t 4\pi G\rho^{\rm bary} \sim O(\nabla^2\Phi)$. Thus, one estimates that $\eta \sim O(\frac{\dot\Phi}{c_0^2})$ if purely sourced by baryon heat flux.

A more rigorous estimation including the scalar field should use

$$\left[\partial_t^2 + \frac{\dot\lambda}{\lambda}\partial_t - c_0^2\nabla^2 - c_0^2(\nabla^2\xi) - c_0^2(\partial_i\xi)\partial_i\right]\eta = \Delta, \tag{C20}$$

where the source term $\Delta \equiv \dot{æ}_i\partial_i(\partial_t \ln\lambda) - \partial_i\frac{\tilde S_i}{2\lambda} \sim [O(\Phi)T_{\rm orb}^{-1} + O(N^{-2})\nu]|\mathbf{k}|^2$ crudely, where we approximate $8\pi G(\partial_i\lambda)\rho^{\rm bary}v_i \sim 8\pi G\partial_i q_i^{\rm bary} \sim O(8\pi G\dot\rho) \sim O(\nabla^2\Phi/T_{\rm orb}) \sim |\mathbf{k}|^2\Phi/T_{\rm orb}$, where $T_{\rm orb}$ = [baryon orbit crossing time]. We also approximate $\dot{æ}_i \sim \partial_i f \sim \partial_i \ln\lambda \sim N^{-2}\mathbf{k}$, which is comparable to $\partial_i\Phi \sim \Phi\mathbf{k}$ if $N^{-1} \sim 10^{-4} > |\Phi|$. Using $\partial_t \ln\lambda \sim \nu\ln\lambda$ with $\nu \sim N/T_{\rm orb}$ being the frequency of the scalar field oscillation, and replacing the partial derivative ∂_i with the wave vector $\mathbf{k}$, we can further estimate

$$\eta = -\partial_j æ_j \sim \tilde f/T_{\rm orb}, \tag{C21}$$

where $\tilde f$ is a fudge factor, estimated by $\tilde f \sim \frac{\Delta}{(\nu^2 - c_0^2|\mathbf{k}|^2)} \sim \frac{O(\sigma^2)+O(N^{-1})}{|N^2\sigma^2 - c_0^2|}$, where $\nu/|\mathbf{k}| \sim N\sigma$, and $\sigma \sim \sqrt{|\Phi|}$ is the typical internal velocity scale of the system. Hence, for most astronomical systems with $\sqrt{|\Phi|} \sim \sigma < \frac{c_0}{N} \sim \frac{300}{N} \times 1000\,{\rm km\,s^{-1}}$, we estimate $1 > \tilde f > \left[\frac{1}{N}, \left(\frac{\sigma}{c_0}\right)^2\right]_{\rm max} \sim \frac{1}{N}$. In short, $\dot{æ}_i$ is typically comparable to the geodesic acceleration, and the damping scale η^{-1} is typically (somewhat longer than) the orbital dynamical timescale in the absence of the Hubble expansion, which tends to damp any evolution of the vector field and the scalar field.

APPENDIX D

SOLAR SYSTEM CONDITIONS FOR OUR CHOICE OF $F(\varphi^2)$

There are several ways to prevent the vector field Z^a from reaching the zero norm, or changing signs of its norm. The easiest is perhaps the replacement,

$$Z^a = \varphi Æ^a \to Z^a = \sqrt{\lambda_\pm}Æ^a = \sqrt{\varphi^2 + \epsilon^2}Æ^a \tag{D1}$$

with a small ϵ.

There are also other ways, keeping the definition $Z^a = \varphi Æ^a$. The norm φ^2 can be prevented from reaching zero because the function $\frac{1}{9}F$ is bounded by infinite high-potential barriers $\epsilon^4\varphi^{-2}$ at $\varphi \to +0$ and φ^2 at $\varphi \to \infty$ with a global minimum $\frac{1}{9}F = -1$ at $\varphi = \epsilon$, which attracts φ with a diverging restoring force

$$-F' \to -(\varphi^2 - \epsilon^2)^{-1/2} \quad \text{if } \varphi^2 > \epsilon^2 \tag{D2}$$

$$\to (\varphi^2 - \epsilon^2)^{-1/2} \text{if } \varphi^2 < \epsilon^2. \tag{D3}$$

So, our dynamical norm vector field Z^a can be approximated as Jacobson's fixed norm field with a small positive norm ϵ^2 in the regime where gravity is high, such as the solar system; the effective Newtonian gravitational constant $G_N = G/(1-\epsilon^2)$, which is virtually the same as the bare G here for a small ϵ^2. Note that Jacobson's fixed norm theory is also the limiting case for a wide class of time-like fields $Æ^a$ with a simple algebraic potential function

$$F(\varphi^2) = [|f|^{1/k} - 1]^3 B_0, \qquad f \equiv \frac{\varphi^n - \epsilon^{2n}\varphi^{-n}}{1 - \epsilon^{2n}}, \tag{D4}$$

where the minimum is at $\phi = \epsilon$, and $B_0 = \frac{1}{3}\left(\frac{2k}{n}\right)^3$ for an appropriate normalization at $\phi = 1$, and $0 < n < \infty$; for example, a model with $k = 3$, $n = 2$ would have $\Lambda_0 F = 9\Lambda_0\left[\left(\frac{\varphi^2 - \epsilon^4\varphi^{-2}}{1-\epsilon^4}\right)^{1/3} - 1\right]$. This would create a potential in cosmology, $9\Lambda_0$ above the minimal $\varphi = \epsilon$ at the solar system. In general, our models have a characteristic gap energy density comparable to the observed vacuum energy density of the universe, and the random kinetic energy density of DM in galaxies. The function $\Lambda_0 F$ is built with the properties to explain both MOND and DE effects.

APPENDIX E

DECOMPOSITION IN THE 3+1 FORMULATION

The main idea of 3 + 1 decomposition is to make spacetime splits of physical quantities of any field (baryon or Aether) with respect to the four-velocity $u^a = g^{ab}u_b$ of an observer. The projection tensor h_{ab} is defined as $h_{ab} = g_{ab} - u_a u_b$ and can be used to obtain covariant tensors perpendicular to u. For example, the covariant spatial derivative $\hat\nabla$ of a tensor field $T_{d\cdots}^{b\cdots}$ is defined as

$$\hat\nabla^a T_{d\cdots}^{b\cdots} \equiv h_j^b \cdots\ h_d^r \cdots h_i^a \nabla^i T_{r\cdots}^{j\cdots}, \quad h_b^a = \delta_b^a - u^a u_b. \tag{E1}$$

The energy–momentum tensor of baryon or Aether and covariant derivative of the *geodesic* four-velocity are decomposed respectively as

$$T_{ab} = \pi_{ab} + 2q_{(a}u_{b)} + \rho u_a u_b - p h_{ab}, \tag{E2}$$

$$\nabla_a u_b = \sigma_{ab} + \varpi_{ab} + \frac{1}{3}\theta h_{ab} + u_a A_b. \tag{E3}$$

In the above, π_{ab} is the projected symmetric trace-free (PSTF) anisotropic stress, q_a is the vector heat flux vector, p is the isotropic pressure, σ_{ab} is the PSTF shear tensor, $\varpi_{ab} = \hat\nabla_{[a}u_{b]}$ is the vorticity, and $\theta = \nabla^c u_c \approx \frac{3}{2}\partial_t \ln a^2(1+2\Psi) \approx 3H + 3\dot\Psi$ is the expansion scalar, which is not negligible even in the absence of Hubble expansion, where the cosmic scale factor $a(t) = 1$. Also note that $A_b = \dot u_b$ is the acceleration; the overdot denotes the time derivative expressed as $\dot\phi = u^a\nabla_a\phi = \partial_t\phi$, brackets mean anti-symmetrization, and parentheses symmetrization. The four-velocity normalization is chosen to be $u^a u_a = 1$. The quantities π_{ab}, q_a, ρ, p are referred to as dynamical quantities and σ_{ab}, ϖ_{ab}, θ, A_a as kinematical quantities. Note that the dynamical quantities can be obtained from the energy–momentum tensor T_{ab} through the relations

$$\begin{aligned} \rho &= T_{ab}u^a u^b, \\ p &= -\frac{1}{3}h^{ab}T_{ab}, \\ q_a &= h_a^d u^c T_{cd}, \\ \pi_{ab} &= h_a^c h_b^d T_{cd} + p h_{ab}. \end{aligned} \tag{E4}$$

Note that $\hat{\nabla}^a = h^{ab}\nabla_b = a^{-2}(0, -\partial_1, -\partial_2, -\partial_3) = -a^{-2}\hat{\nabla}_a$ is the spatial derivative, where 1, 2, 3 stand for three comoving coordinates, and $\hat{\nabla}^2 = \hat{\nabla}^a\hat{\nabla}_a = -a^{-2}(\partial_1^2 + \partial_2^2 + \partial_3^2) = -\nabla^2 = -(\partial_x^2 + \partial_y^2 + \partial_z^2)$, where ∇^2 is the usual Laplacian in proper coordinates (x, y, z).

Note that kinematical quantities like θ also satisfy constraint equations like

$$8\pi G q_a = -\frac{2}{3}\hat{\nabla}_a\theta + \hat{\nabla}^b\sigma_{ab} + \hat{\nabla}^b\varpi_{ab} \tag{E5}$$

(Li et al. 2008) and propagation equations like

$$4\pi G(\rho + 3p) = -\dot{\theta} - \frac{\theta^2}{3} + \hat{\nabla}^a A_a. \tag{E6}$$

All 3+1 formulated physical quantities are expressed in terms of $Æ_a = (0, æ_1, æ_2, æ_3)$, and the kinematical quantities, such as the local volume expansion rate $\theta = \nabla_a u^a = 3H + 3\dot{\Psi}$ and the local acceleration A_i; hence there is no need to evolve the geodesic equation of u^a explicitly.

REFERENCES

Angus, G. W., & McGaugh, S. 2008, MNRAS, 383, 417
Angus, G. W., Shan, H., Zhao, H., & Famaey, B. 2007, ApJ, 654, L13
Baumgardt, H., Grebel, E. K., & Kroupa, P. 2005, MNRAS, 359, L1
Bekenstein, J. 1988, Phys. Lett. B, 202, 497
Bekenstein, J. 2004, Phys. Rev. D, 70, 3509
Bekenstein, J., & Milgrom, M. 1984, ApJ, 286, 7
Bekenstein, J., & Sanders, R. 1994, ApJ, 429, 480
Bento, M. C., Bertolami, O., & Sen, A. A. 2002, Phys. Rev. D, 66, 043507
Brans, C. H., & Dicke, R. H. 1961, Phys. Rep., 124, 925
Bruneton, J.-P., & Esposito-Farese, G. 2007, Phys. Rev. D, 76, 124012
Carroll, S., & Lim, E. 2004, Phys. Rev. D, 70, 123525
Chiba, T. 2003, Phys. Lett. B, 575, 1
Ciotti, L., & Binney, J. 2004, MNRAS, 351, 285
Debattista, V. P., & Sellwood, J. A. 1998, ApJ, 493, L5
Dodelson, S., & Liguori, M. 2006, Phys. Rev. Lett., 97, 1301
Ferreira, P., Gripaios, B. M., Saffari, R., & Zlosnik, T. G. 2007, Phys. Rev. D, 75, 044014
Foster, B. J., & Jacobson, T. 2006, Phys. Rev. D, 73, 064015
Graesser, M. L., Jenkins, A., & Wise, M. B. 2005, Phys. Lett. B, 613, 5
Halle, A., Zhao, H., & Li, B. 2008, ApJS, 177, 1
Jacobson, T., & Mattingly, D. 2001, Phys. Rev. D, 64, 024028
Kanno, S., & Soda, J. 2006, Phys. Rev. D, 74, f3505
Kostelecky, V. A., & Samuel, S. 1989, Phys. Rev. D, 39, 683
Li, B., Mota, D. F., & Barrow, J. D. 2008, Phys. Rev. D, 77, 024032
Lim, E. A. 2005, Phys. Rev. D, 71, 063504
Nipoti, C., Londrillo, P., Binney, J., & Ciotti, L. 2008, MNRAS, 386, 2194
Peebles, J., & Vilenkin, A. 1999, Phys. Rev. D, 59, 063505
Sanchez-Salcedo, F. J., & Hernandez, X. 2007, ApJ, 667, 878
Sanders, R. H. 1997, ApJ, 480, 492
Sanders, R. H. 2005, MNRAS, 363, 459
Skordis, C., Mota, D., Ferreira, P., & Boehm, C. 2006, Phys. Rev. Lett., 96, 011301
Tiret, O., & Combes, F. 2008, A&A, 483, 719
Will, C. M. 1993, Theory and Experiment in Gravitational Physics (Cambridge: Cambridge Univ. Press), 126
Zhao, H. S. 2005, A&A, 444, L25
Zhao, H. S. 2007, ApJ, 671, L1
Zhao, H. S. 2008a, arXiv:0805.4046
Zhao, H. S. 2008b, J. Phys.: Conf. Ser., 140, 012002
Zlosnik, T., Ferreira, P., & Starkman, G. 2007, Phys. Rev. D, 75, d4017
Zuntz, J. A., Ferreira, P., & Zlosnik, T. 2008, Phys. Rev. Lett., 101, 261102